Estimating negative binomial parameters from occurrence data with detection times

Wen-Han Hwang[1], Richard Huggins[2], and Jakub Stoklosa[3]

[1] Institute of Statistics, National Chung Hsing University, Taiwan.

[2]Department of Mathematics and Statistics, The University of Melbourne, Australia.

[3]School of Mathematics and Statistics and Evolution & Ecology Research Centre,

The University of New South Wales, Australia.

May 24, 2016

Abstract

The negative binomial distribution is a common model for the analysis of count data in biology and ecology. In many applications, we may not observe the complete frequency count in a quadrat but only that a species occurred in the quadrat. If only occurrence data are available then the two parameters of the negative binomial parameters, the aggregation index and the mean, are not identifiable. This can be overcome by data augmentation or through modelling the dependence between quadrat occupancies. Here we propose to record the (first) detection time while collecting occurrence data in a quadrat. We show that under what we call proportionate sampling, where the time to survey a region is proportional to the area of the region, that both negative binomial parameters are estimable. When the mean parameter is larger than two, our proposed approach is more efficient than the data augmentation method developed by Solow & Smith (2010, *Amer. Nat.* **176**, 96–98), and in general is cheaper to conduct. We also

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/bimj.201500239

investigate the effect of mis-identification when collecting negative binomially-distributed data, and conclude that, in general, the effect can be simply adjusted for provided that the mean and variance of mis-identification probabilities are known. The results are demonstrated in a simulation study and illustrated in several real examples.

Key words: Aggregation index; Cost Analysis; Mis-identification; Negative binomial distribution, Presence-absence data.

1 Introduction

Population ecology data are often collected through quadrat sampling where some plots of a standard area are randomly selected and a survey is conducted within each plot (Manly & Navarro Alberto, 2015). In biology and ecology applications, it is common to count the frequency (that is, the number of individuals) for a particular species in a sampled quadrat. Many distributions can be used to model these count data of which the most popular methods are using a Poisson or negative binomial distribution (Cameron & Trivedi, 1998; Ver Hoef and Boveng, 2007; Winkelmann, 2008); the later is more flexible (though more complicated) as it includes the former one as a limiting case under some conditions. Spatial models based on the negative binomial distribution are not new. For example, Gregoire (1983) and Diggle & Milne (1983) considered the negative binomial distribution for quadrat counts, and Alexander *et al.* (2000) considered a Bayesian approach to inference on spatial negative binomial models and applied this to model parasite counts. Our proposed model differs from these in that we only have partial observation over quadrats. However, our approach is related to mechanism II' of Diggle & Milne (1983).

In general, occupancy models only observe whether a quadrat is occupied or not, and not the actual count. From a management perspective, this can greatly reduce the cost of sampling, as a quadrat is only examined until the first individual is detected and only

exhaustively examined if no individual is detected. More recently Solow & Smith (2010) proposed surveying a quadrat until the second individual is detected or the quadrat has been completely surveyed so that quadrats are either unoccupied, occupied by exactly one individual, or occupied by more than one individual. Importantly, there is an increase in the survey cost when considering this approach.

There are two parameters in a negative binomial distribution, these are: the shape parameter (κ) and a mean parameter (μ). The shape parameter κ is related to the degree of overdispersion and often serves as a measure (index) describing the degree of aggregation (clustering) for the distribution of a species (Pielou, 1977). Aggregation of the species increases when $\kappa \to 0$, and vice versa as $\kappa \to \infty$. It is therefore of key interest to ecologists and biologists to estimate and understand the shape parameter. Hereafter we refer to κ as the *aggregation index*. Note that, when $\kappa \to \infty$ the negative binomial distribution converges to a Poisson distribution. However, the parameters are un-identifiable if only occurrence data are given, since there is only one sufficient statistic "occupancy" (that is, the number of occurrences/sample size), see Conlisk *et al.* (2007). As a result, in order to solve this difficulty, it requires additional information besides the occurrence data. Conlisk *et al.* (2009) considered a regression model using a notion of cross cell occupancy clustering. As noted above, Solow & Smith (2010) needed to identify the frequency of single occupancies – we discuss this approach in greater detail in Section 2. Yin & He (2014) considered occurrence map data and developed a model that takes into account spatial autocorrelation. Hwang & Huggins (2015) also considered the occurrence map and assumed dependence between neighbouring quadrats which they modelled as a multivariate negative binomial distribution.

To retain the cheaper cost of a sampling scheme that surveys each quadrat until the first individual is detected, we propose to additionally record the time to the first detection (or hitting time) in each quadrat. We illustrate this in Figure 1 where we show an example of the proposed data collection scheme and compare this with the Solow & Smith (2010) sampling method. We show that under this scheme, if the time to survey a region is proportional to its area then both of the parameters in the negative binomial

distribution are estimable without further assumptions or additional data. We also conduct a cost analysis and investigate the effect of mis-identification (detection error), and conclude that, in general there is a simple adjustment provided that the mean and variance of mis-identification probabilities are known. To the best our knowledge, there is currently no literature that considers this type of mis-identification effect.

In Section 2, we develop a model when using first detection times of occurence and define proportionate sampling. In Section 3, we derive the expected survey times and costs of collecting the data for both the Solow & Smith (2010) approach and our proposed method. We then consider a model that allows mis-identification in Section 4. In Section 5, we conduct a simulation study to evaluate the performance of the proposed method, and in Section 6, we apply the method to real data examples. We conclude the study with some discussion in Section 7.

2 Model and Method

The probability function of a negative binomial distribution is

$$f_{\boldsymbol{\theta}}(x) = \frac{\Gamma(\kappa + x)}{\Gamma(\kappa)x!} \frac{\kappa^{\kappa}\mu^{x}}{(\kappa + \mu)^{x+\kappa}}, x = 0, 1, 2, \ldots, \tag{1}$$

where $\boldsymbol{\theta} = (\kappa, \mu)$, with κ being the aggregation index, μ being the mean and $\Gamma(\cdot)$ is the usual gamma function. We write $X \sim NB(\kappa, \mu)$ to denote the random variable X which has a negative binomial distribution with probability function (1). Then $E(X) = \mu$ and $Var(X) = \mu + \mu^2/\kappa$ where κ controls the degree of overdispersion relative to the variance of a Poisson count.

Consider a random sample $X_1, \ldots, X_n$ from a $NB(\kappa, \mu)$, representing the number of individuals occupying the n quadrats. In occupancy models we only have the binary observations Y_i, $i = 1, \ldots, n$ where $Y_i = I(X_i > 0)$ takes the value zero for an absence and 1 otherwise. Then the Y_i, $i = 1, \ldots, n$ are independent Bernoulli random variables. We wish to use these data to estimate the vector of parameters $\boldsymbol{\theta}$. Let p_+ denote the

probability of presence, i.e., $p_+ = P(Y_i = 1) = 1 - (1 + \mu/\kappa)^{-\kappa}$. The maximum likelihood estimate (MLE) of p_+ is the occupancy rate m/n where $m = \sum_i Y_i$ (the number of presences). Nevertheless, $\boldsymbol{\theta}$ is not identifiable since m is the single sufficient statistic using the distribution of the Y_i.

To resolve this, in addition to the occurrence data, Solow & Smith (2010) recommended that each presence observation can be further identified as two cases: "singleton" and "two or more", where the singleton means that there is exactly a single individual from a complete observation of the quadrat. Let $m_0 = n - m$, m_1 be the number of singletons and $m_2 = m - m_1$. It follows that (m_0, m_1, m_2) has a multinomial distribution with corresponding probabilities $(p_0(\boldsymbol{\theta}), p_1(\boldsymbol{\theta}), p_2(\boldsymbol{\theta}))$ where $p_j(\boldsymbol{\theta}) = f_{\boldsymbol{\theta}}(j)$ for $j = 0, 1$ and $p_2(\boldsymbol{\theta}) = 1 - p_0(\boldsymbol{\theta}) - p_1(\boldsymbol{\theta})$. We then proceed as usual by maximizing the following log-likelihood function to obtain the MLE

$$L_1(\boldsymbol{\theta}) = \sum_{j=0}^{2} m_j \log \{p_j(\boldsymbol{\theta})\}.$$

This method requires surveying a quadrat until the second individual is detected or until the quadrat has been completely surveyed (i.e., either zero or one individual is detected in the quadrat).

Let t_i, $i = 1, \ldots, n$ be the first detection time for the survey in the ith quadrat. To model the distribution of t_i we need to consider how the search is conducted. Recall that a negative binomial distribution is also equivalent to a gamma-Poisson mixture distribution, i.e., $X \sim NB(\kappa, \mu)$ can be viewed as a mixture of Poisson distributions where the mixing distribution of the Poisson rate (say λ) follows a gamma distribution, $Gamma(\kappa, \mu/\kappa)$. Here $Gamma(\kappa, \mu/\kappa)$ denotes a gamma distribution with mean μ and variance μ^2/κ. Accordingly, we suppose that given λ occurrences over the quadrat form a spatial Poisson process with rate λ. It follows that the number of individuals within an area of size b (of the quadrat) has a Poisson distribution with mean $b\lambda$. Now, let a_j be the area that needs to be surveyed to find the jth individual, then $P(a_1 > b \mid \lambda) = P(\text{zero occurrence within an area } b \mid \lambda) = e^{-b\lambda}$ which implies that the conditional density of a_1 given λ is exponential with mean $1/\lambda$ (i.e.,

we have $\lambda \sim Gamma(1, 1/\lambda)$). Following a similar argument as above and using the independence increment property of a Poisson process, the conditional density of a_j given λ is $Gamma(j, 1/\lambda)$.

Thus, if the first detection time is $t = \alpha a_1$ for some $\alpha > 0$, then the conditional density of t given λ is $Gamma(1, \alpha/\lambda)$, and the unconditional density of t is

$$g(t) = \int_0^\infty \frac{\lambda}{\alpha} e^{-\frac{\lambda}{\alpha}t} \frac{\lambda^{\kappa-1}}{\Gamma(\kappa)} \frac{\kappa^\kappa}{\mu^\kappa} e^{-\frac{\kappa}{\mu}\lambda} \mathrm{d}\lambda = \frac{\mu\alpha^\kappa \kappa^{\kappa+1}}{(\kappa\alpha + \mu t)^{\kappa+1}}.$$

Similarly, the density of the time to the second detection is

$$h(t) = \int_0^\infty t\frac{\lambda^2}{\alpha^2} e^{-\frac{\lambda}{\alpha}t} \frac{\lambda^{\kappa-1}}{\Gamma(\kappa)} \frac{\kappa^\kappa}{\mu^\kappa} e^{-\frac{\kappa}{\mu}\lambda} \mathrm{d}\lambda = \frac{t\mu^2\alpha^{\kappa-1}\kappa^{\kappa+1}(\kappa+1)}{(\kappa\alpha + \mu t)^{\kappa+21}}.$$

These expressions hold for a search starting at any point in the quadrat, as long as the time to survey a region is proportional to the area of the region. We call this "**proportionate sampling**". This is often a reasonable assumption in practice but may be violated, for example, if the terrain is not uniform across the sampling area. It is more convenient to change the time scale by dividing the time by α, doing this permits the following results:

Proposition 1. *Assume that $X \sim NB(\kappa, \mu)$, then under proportionate sampling, the probability density function of the first detection time is*

$$g_{\boldsymbol{\theta}}(t) = \frac{\mu\kappa^{\kappa+1}}{(\kappa + \mu t)^{\kappa+1}}, t > 0,$$

and that of the time to the second detection is

$$h_{\boldsymbol{\theta}}(t) = \frac{t\mu^2(\kappa+1)\kappa^{\kappa+1}}{(\kappa + \mu t)^{\kappa+2}}, t > 0.$$

The proof of Proposition 1 follows from the aforementioned $g(t)$ and $h(t)$ by taking $\alpha = 1$. Without loss of generality, we assume that the area of each quadrat is one. Then the (re-scaled) time required to completely survey each quadrat is also one and the detection time is equal to percentage of surveyed area as shown in Figure 1.

Therefore, $t_i > 1$ if and only if $Y_i = 0$. Suppose that $t_i < 1$ for $i = 1, \ldots, m$ and $t_i = 1$ otherwise. Then the log-likelihood function is

$$L_2(\boldsymbol{\theta}) = \sum_{i=1}^{m} \log g_{\boldsymbol{\theta}}(t_i) + (n - m) \log \{p_0(\boldsymbol{\theta})\} . \tag{2}$$

As usual, we can use (2) to obtain the MLE of $\boldsymbol{\theta}$ and the observed Fisher information to calculate standard errors. In Figure 2, we compare the efficiency of $L_1(\boldsymbol{\theta})$ and $L_2(\boldsymbol{\theta})$ by showing the ratio $det\{I_2(\boldsymbol{\theta})\}/det\{I_1(\boldsymbol{\theta})\}$ where det denotes the determinant of a matrix, and $I_j(\boldsymbol{\theta})$ is the expected Fisher information matrix of $L_j(\boldsymbol{\theta})$ for $j = 1, 2$. More specifically, this demonstrates that our approach is more efficient than that of Solow & Smith (2010) when $\mu > 2$.

Finally we note that the assumption of proportionate sampling is useful when converting the method to a temporal scale. In practice, we only have to record the percentage of surveyed area in a quadrat for detecting the first individual (e.g., see Figure 1). Moreover, Proposition 1 holds not only for quadrat sampling over space but also for other sampling types which are designed/collected within a standard unit of time, volume, etc., for example, samples that are collected on bird counts within, say, 5 minutes or the number of micro-organisms contained in measures of volume of soil or water.

3 Cost Analysis

As well as differences in efficiency there is a difference in cost between our approach and that of Solow & Smith (2010), in fact any method that requires augmented data. Consider a quadrat containing X individuals. Let T_1 and T_2 be the times until the detection of the first and second individuals, respectively, where these are one if the individual was not detected. We compare the expectations of T_1 and T_2 below. The proof is given in the Appendix.

Proposition 2. *Assume that* $X \sim NB(\kappa, \mu)$ *and sampling is proportionate, then* $E(T_1) =$

$E\{1/(X+1)\}$, $E(T_2) = 2E(T_1) - P(X=0)$ *and*

$$E\left(\frac{1}{X+1}\right) = \begin{cases} \dfrac{\kappa}{\mu(\kappa-1)}\left\{1 - \dfrac{1}{(1+\frac{\mu}{\kappa})^{\kappa-1}}\right\}, & \kappa \neq 1, \\ \dfrac{\log(1+\mu)}{\mu}, & \kappa = 1. \end{cases}$$

Moreover,

$$E(T_2 - T_1) = \begin{cases} \dfrac{\kappa}{\mu(\kappa-1)}\left\{1 - \left(\dfrac{\kappa}{\kappa+\mu}\right)^{\kappa-1}\right\} - \left(\dfrac{\kappa}{\kappa+\mu}\right)^{\kappa}, & \kappa \neq 1, \\ \dfrac{\log(1+\mu)}{\mu} - \dfrac{1}{1+\mu}, & \kappa = 1. \end{cases}$$

This proposition allows us to compare the expected cost of our proposed scheme (where we sample to the first detection and record the time) with the scheme of Solow & Smith (2010) that requires sampling until the second detection. In Web Figure 1 we plot the percentage expected cost saving $1 - E(T_1)$ compared with exhaustive sampling, and $E(T_2) - E(T_1)$ when compared with the proposal of Solow & Smith (2010). In the first case, the savings increase with μ. In the second case the savings are maximised for μ around 2 and can be substantial, although this is somewhat expected. For small μ both surveys are close to exhaustive if not exhaustive, and for larger μ, the expected time between detections becomes small.

Proposition 3. *Assume that* $X \sim NB(\kappa, \mu)$ *and sampling is proportionate, then* $Var(T_1) = 2E[1/\{(X+2)(X+1)\}] - E^2(T_1)$ *and* $Var(T_2) = 6E[1/\{(X+2)(X+1)\}] - E^2(T_2) - 2P(X=0)$. *Moreover, we have* $Cov(T_1, T_2) = 3E[1/\{(X+2)(X+1)\}] - E(T_1)E(T_2) - P(X=0)/2$.

The proof is also given in the Appendix. A further detailed calculation leads to following result. If $\kappa = 1$, then $E[1/\{(X+2)(X+1)\}] = \left\{-\log(1+\mu)/\mu^2 + 1/\mu\right\}$, and if $\kappa = 2$, then $E[1/\{(X+2)(X+1)\}] = \left[4\log\left(1+\mu/2\right)/\mu^2 - 4/\{\mu(2+\mu)\}\right]$. For other cases of κ,

$$E\left\{\frac{1}{(X+2)(X+1)}\right\} = \frac{\kappa^2}{\mu^2(\kappa-1)(\kappa-2)}\left\{1 - \frac{\kappa^{\kappa-2}}{(\kappa+\mu)^{\kappa-2}} - \frac{(\kappa-2)\mu\kappa^{\kappa-2}}{(\kappa+\mu)^{\kappa-1}}\right\}.$$

4 Mis-identification of Detections

It may be possible that an individual is mis-identified during sampling. In practice, there are two types of mis-identification errors, namely: "false positive" and "false negative", here we only consider the latter type. That is, an individual of the target species say, species A is mis-identified as belonging to another, say, species B. Suppose that c is the probability of correct identification of an individual. That is, c is the conditional probability that an individual that has been discovered is then correctly identified. We assume that detections are made independently over all individuals and that individuals of another species cannot be mis-identified as the species of interest.

Let $X \sim NB(\kappa, \mu)$ and $X' \leq X$ be the number of identified individuals. Under the assumption of independent detections, we have $P(X' = x') = \sum_{x=x'}^{\infty} \binom{x}{x'} c^{x'}(1-c)^{x-x'} P(X = x)$ for any non-negative integer x' which gives $X' \sim NB(\kappa, c\mu)$. Thus, by Proposition 1, the distribution of the first detection time now becomes

$$g'_{\boldsymbol{\theta}}(t) = \frac{c\mu\kappa^{\kappa+1}}{(\kappa + c\mu t)^{\kappa+1}}, t > 0. \tag{3}$$

Using the results of Section 2 we can find MLE's for κ and $c\mu$. However, it is clear from the form of the density (3) that c and μ are not identifiable, and we can only estimate μ if c is known or can be estimated from ancillary data. Importantly though, κ is identifiable.

In practice, the correct detection probability c is likely to be a random variable. For example, some data collectors may be well trained but others may lack experience in the required sampling scheme. Suppose that the ith quadrat is surveyed with correct identification probability C_i where $E(C_i) = c$ and $Var(C_i) = \sigma_c^2$. Hereafter, we assume that both c and σ_c^2 are known – although in practice both values are either given or can be estimated from some validation sample, and/or replicated observations. Notice that the parameter κ can be underestimated even if the mean c is known. To see this, suppose that $X'_i \sim NB(\kappa, C_i\mu), i = 1, \ldots, n$ when C_i is given. The moment estimators of μ and κ would be $\tilde{\mu} = \overline{X'}/c$ and $\tilde{\kappa} = \overline{X'}^2/(S^2_{x'} - \overline{X'})$, respectively. By the law of

large numbers, we have

$$\begin{aligned} \overline{X'} &\longrightarrow c\mu, \\ \overline{X'^2} &\longrightarrow (c^2+\sigma_c^2)\mu^2\left(1+\frac{1}{\kappa}\right)+c\mu. \end{aligned}$$

Therefore, if we did not take the variation of C_i into account, then $\tilde{\kappa} \to \kappa/\{1+(1+\kappa)\sigma_c^2/c^2\}$ so that $\tilde{\kappa}$ would underestimate the true κ with probability 1 as the sample size n increases to infinity. This is an attenuation effect commonly seen in the context of measurement error modelling (Carroll *et al.*, 2006). Although not theoretically confirmed, we conjecture that the MLEs may also exhibit this behaviour. A simple way to correct for the estimate of κ is by multiplying by the following adjustment quantity $1+(1+\hat{\kappa}_0)\sigma_c^2/c^2$ where $\hat{\kappa}_0$ is a nàive estimate of κ that ignores the effect of the variation of C_i. Note that, in the worst case, the attenuation coefficient $1/\{1+(1+\kappa)\sigma_c^2/c^2\}$ can be $1/(2+\kappa)$. But in general applications this may be minor, e.g., if $c \geq 80\%$ and $\sigma_c \leq 0.1$ then the number falls in the range $(64/(65+\kappa), 1)$ so that it may not be necessary to do the adjustment if κ is not too large. We confirm this in a simulation study in Section 5.

A formal estimating procedure is to maximize the marginal likelihood function

$$\sum_{i=1}^{n} \log\left\{\int \mathcal{L}_i \underline{L_i}(\boldsymbol{\theta} \mid c)\rho(c)\mathrm{d}c\right\}, \tag{4}$$

where $\mathcal{L}_i \underline{L_i}(\boldsymbol{\theta} \mid c)$ is the likelihood function of the ith observation, e.g., $NB(\kappa, c\mu)$ or equation (3), and $\rho(c)$ is the probability density function of C_i. Here a parametric assumption for $\rho(\cdot)$ should be specified, along with some knowledge of c and σ_c^2. In general, we may use numerical integration to calculate the marginal likelihood for implementing this procedure. However, in some cases, the marginal likelihood function may have an explicit form. As an example, consider the probability density function (3)

and assume that $C_i \sim U(a,b)$, then the marginal probability density function of t_i is

$$\int_a^b \frac{c\mu\kappa^{\kappa+1}}{(b-a)(\kappa+c\mu t_i)^{\kappa+1}}\mathrm{d}c = -\frac{c\kappa^{\kappa}}{t_i(b-a)(\kappa+c\mu t_i)^{\kappa}}\Big|_a^b + \int_a^b \frac{\kappa^{\kappa}}{t_i(b-a)(\kappa+c\mu t_i)^{\kappa}}\mathrm{d}c$$

$$= \begin{cases} \frac{1}{(b-a)}\left[\frac{\kappa^{\kappa}}{t_i}\left\{\frac{a}{(\kappa+a\mu t_i)^{\kappa}} - \frac{b}{(\kappa+b\mu t_i)^{\kappa}}\right\}\right. \\ \qquad \left. + \frac{\kappa^{\kappa}}{(\kappa-1)\mu t_i^2}\left\{\frac{1}{(\kappa+a\mu t_i)^{\kappa-1}} - \frac{1}{(\kappa+b\mu t_i)^{\kappa-1}}\right\}\right], & \kappa \neq 1, \\ \\ \frac{1}{(b-a)}\left[\frac{1}{t_i}\left\{\frac{a}{(1+a\mu t_i)} - \frac{b}{(1+b\mu t_i)}\right\}\right. \\ \qquad \left. + \frac{1}{\mu t_i^2}\left\{\log\left(\frac{1+b\mu t_i}{1+a\mu t_i}\right)\right\}\right], & \kappa = 1. \end{cases}$$

5 Simulation Studies

We carried out several simulation studies to evaluate the finite sample performance for the methods presented in Sections 2–4. In the first simulation study we primarily focused on the performance of model parameters, and in the second simulation study we examined the effects of mis-identification of detection.

5.1 Simulation study 1

We considered the following cases: $\kappa = 0.5, 1, 5$, $\mu = 0.5, 2, 5$ and sample sizes $n = 100, 200, 500$. For each combination, we then generated count data $X_i, i = 1, \ldots, n$ from $NB(\kappa, \mu)$ for each value of κ and μ as above. First detection times t_i were generated from the conditional probability density $g_{\boldsymbol{\theta}}(t)/\int_0^1 g_{\boldsymbol{\theta}}(u)du = g_{\boldsymbol{\theta}}(t)/\{1-p_0(\boldsymbol{\theta})\}, 0 < t < 1$ for those $X_i > 0$, and $t_i = 1$ when $X_i = 0$. Note that these detection times can be obtained by using an inverse probability integral transformation. For each scenario, 1000 samples were generated. For each sample, we computed estimates and standard errors using maximum likelihood estimation (i.e., fitting the negative binomial model) from the complete uncensored data (which we abbreviate as "Comp", or

simply refer to as the complete MLE throughout), using the Solow & Smith (2010) approach which requires the frequency data on $0, 1$ and $2+$ (which we abbreviate to "S&S" throughout), and the proposed method which uses occurrence data with detection times (which we abbreviate to "DT" throughout). Note that the complete MLEs were only calculated here to be used as a baseline reference method, our main objective was to compare the performance of the S&S and DT methods.

In the simulated data sets, we occasionally obtained extreme outliers which skewed the overall results (see Section 7 for a further discussion). Therefore, instead of reporting the usual average and standard deviation, we reported the median of the estimate (Med), the rescaled median absolute deviation (MAD), the median of the estimated standard errors (M.SE), and the sample coverage percentage (CP) of trials in which 95% Wald-type confidence intervals covered the true parameters. For each method, we excluded some samples due to the optimization algorithm failing to converge to a solution, however, non-convergence only occurred occasionally in these studies. Specifically, the non-convergence percentages for Comp and DT methods were less than 0.5% in all cases, but for the S&S approach this occurred about $2-4\%$ of the time when $n \leq 200$ and $\kappa \geq 1$. Note that we used the `optim` function in `R` (`R` Development Core Team, 2016) to calculate the estimates. We also presented boxplots which gave a graphical representation of the results and the extent of variability in estimating model parameters.

In Figures 3–6 we give boxplots for the estimates of κ (top) and μ (bottom) when: (i) $\kappa = 0.5, \mu = 0.5$, (ii) $\kappa = 0.5, \mu = 2$, (iii) $\kappa = 1, \mu = 2$ and (iv) $\kappa = 1, \mu = 5$, for each sample size. In the Web Supplementary materials we give boxplots for all other remaining combinations. The expected detection times $E(T_1)$ and $E(T_2)$ are also reported in these tables.

For moderate to large values of μ the DT method performed well, giving little bias and close to nominal 95% CPs when compared with the Comp method, e.g., see Figures 4–6 and Web Tables 2–3. In contrast to the S&S approach, the DT method yielded similar (and in most case better) coverage for both κ and μ when using moderate to large

means μ, regardless of the sample size used. The MADs/M.SEs for the DT method were larger than the Comp method but were smaller than the S&S approach. For small μ and small sample sizes, the DT method slightly underperformed compared to the S&S approach (see Figure 3 and Web Table 1), in particular for small aggregation values (e.g., $\kappa = 5$), however all methods (including the complete MLE case) performed poorly for large κ, as seen in the bottom rows of Web Table 1. Note that, negative binomial parameters are almost unidentifiable when μ is small and/or κ is large, e.g., see Cruyff & Van der Heijden (2014). Naturally, all methods greatly improved when the sample was increased, in fact the DT method outperformed the S&S approach in terms of efficiency and coverage when $\mu \geq 2$ (see Web Tables 2–3). The variability was similar for the S&S approach and the DT method when $\mu \approx 2$ (e.g., see Figure 4 and 5), and the DT method was generally more efficient for larger μ, e.g., see Figure 6.

In summary, under these simulated settings, the DT method can perform poorly for small μ and large κ, in which case the S&S approach may be preferred however for small to moderate values of κ, the proposed DT method would be more beneficial, particularly for large sample sizes.

5.2 Simulation study 2

Following the simulation set up as above, we suppose that the correct identification probabilities were $C_i \sim N(c, \sigma_c^2)$ for all $i = 1, \ldots, n$ and are independent of each other. We set $c = 0.7$ and $\sigma_c = 0.1$ in the simulation study. For each method, we considered two types of adjustment methods: (a) coefficient adjustment (with assumed knowledge on c and σ_c); and (b) a marginal likelihood approach (with assumed knowledge on $C_i \sim N(c, \sigma_c^2)$). Specifically, let $(\hat{\kappa}, \hat{\mu})$ be náive estimates that do not consider the effects of mis-identification. The coefficient adjustment method takes $c^{-1}\hat{\mu}$ and $\{1 + (1 + \hat{\kappa})\sigma_c^2/c^2\}\,\hat{\kappa}$ as the resulting estimates. Moreover, the same adjustments are applied to the associated standard errors. Clearly, the multiplied coefficients are from the moment method as discussed in Section 4. A marginal likelihood approach

was used to compute the MLE by maximizing the likelihood function (4). Due to the normality assumption on C_i, the likelihood function can be easily calculated using Gauss–Hermite quadrature. Thus, we calculated the integral involved in (4) using Gauss–Hermite quadrature with 10 nodes.

In Web Table 4 we give the results for $\mu = 3$, $n = 500$, as well as $\kappa = 0.5, 1$ and 5. We denote the coefficient adjustment methods by Comp_{ca}, $\text{S\&S}_{ca}$, and DT_{ca}; and the marginal likelihood approach by Comp_{ml}, $\text{S\&S}_{ml}$, and DT_{ml}. As seen in Web Table 4, the relative performance among these three methods are similar to those for $n = 500$ in Web Table 3 since now we have the mean $c\mu = 2.1$ which is approximately equal to $\mu = 2$. The simple coefficient adjustment method appeared to work well. We note that the adjustment coefficient is $1+(1+\hat{\kappa})/49$, and so this is rather minor in both cases for $\kappa = 0.5$ and 1. However, when $\kappa = 5$, this type of adjustment made a substantial improvement. The marginal likelihood approach performed satisfactorily in all cases. Nevertheless, although this is a formal procedure and requires the specification of the distribution of C_i, it did not outperform the simple coefficient adjustment in all situations reported here.

We additionally conducted a simulation study with $c = 0.85$ and $\sigma_c = 0.05$ (not reported) and found that the effects of mis-identification were minor (as similarly seen in Section 4), whereas the correction methods yielded more satisfactory results.

6 Examples

We applied our new method to a variety of real data examples consisting of count observations with complete frequencies. In the first case study (Section 6.1), we give examples where the count data did not have first detection time records; thus, in order to fit the proposed models we generated first detection times using the complete data. In the second case study (Section 6.2) we present an example where detection effort was available. For all the examples given below, complete frequencies were available

so we were able to compare our estimates with both the complete MLE and the S&S methods.

6.1 Case study 1: Examples with simulated detection times

To generate detection times, suppose that our data consists of $x_i, i = 1, \ldots, n$ where $x_i > 0$ for $i \leq m$ and zero otherwise. The first detection times t_i can be generated as follows. Clearly, we can define $t_i = 1$ if $x_i = 0$. For a positive observation $x_i > 0$, we view x_i as the number of events from a homogeneous Poisson process in the time interval $[0, 1]$, so the x_i event times are independently and uniformly distributed between 0 and 1 (Daley & Vere-Jones, 2003), thus we generated $U_{ij},\ j = 1, \ldots, x_i$ independent values from the uniform distribution, such that $t_i = \min\{U_{ij}, j = 1, \ldots, x_i\}$. Note that when using the estimators given by Solow & Smith (2010) it is straightforward to calculate MLEs from count data. The time spent by Solow & Smith (2010) is 1 for $x_i = 0, 1$, and for $x_i > 1$ this was generated by the second minimum of $\{U_{ij}, j = 1, \ldots, x_i\}$. For each example data set presented below, we repeated this simulation strategy 1000 times; this allowed us to calculate the sample average and median for κ and μ when using the artificial detection times.

As in Section 5, we use the same abbreviations for each method, and additionally denote DT_a and DT_m as the sample average and median, respectively, which correspond to the DT estimates from the 1000 simulated data sets (the sample average and sample median of the standard error estimates are given in the parentheses). We also give the average (and s.d.) for the detection times for each method from the 1000 simulated data sets.

In Table 1, for each data set presented below, we give the number of zeros (n_0), the sample mean ($\bar{x}$), the sample standard deviation ($sd(x)$), and we test for H_0 : $x \sim NB$ by reporting the p-value. Note that, p-values were evaluated using the Chi-square goodness-of-fit statistic from the `gofstat()` function in the `fitdistrplus` `R`-package (Delignette-Muller & Dutang, 2015). All examples presented below do not

reject the null hypothesis at a significance level of $\alpha = 0.05$, indicating that these count data likely follow the negative binomial distribution.

Abundance data on feline roundworms *Toxocara cati*

In our first example, the data set consists of counts of parasites (feline roundworms) collected on feral cats on Kerguelen Island. Of particular interest was the abundance of feline roundworm parasites *Toxocara cati* found in the digestive tract in feral cats, see Fromont *et al.* (2001) for further details. These data were obtained from the `fitdistrplus` `R`-package. The sample size is $n = 53$.

We give the results in the top part of Table 1. The proposed DT method yielded similar estimates (for κ and μ) compared to the Comp method – although this may be expected since the sample mean $\bar{x} = 8.68$ was quite large, see Simulation Study 1 in Section 5.1. The S&S approach overestimated κ and underestimated μ compared with the Comp method. The average (time) for the first detection time was 0.424(0.022) (the number in the parentheses is the standard deviation from 1000 generated samples). In contrast, the average for the second detection time was 0.584(0.017).

Albatrosses incidental capture data

Our second example data set consists of incidental captures of albatrosses *Phoebastria albatrus* in the sub-Antarctic squid trawl fishery in New Zealand. These data were previously analysed in Hilborn & Mangel (pp. 100, 1997) where the numbers of birds trapped accidentally in nets or trawl gear/cables (in 1990) were of main interest. The sample size is $n = 897$.

Here, the sample size was large but the number of zeros was also very large, which yielded a small sample mean of 0.28. Nevertheless, the proposed DT approach gave very similar estimates for κ and μ compared to the Comp method, with the reported standard errors being slightly larger. The S&S approach gave slightly large estimates

for μ and smaller estimates for κ when compared to the proposed DT method. The average times for the first and second detection times were quite similar, and they were close to 1 (i.e., time for exhausted detection).

Bacteria count data

The third example data set consists of counts of water bacteria colonies (per millilitre) found in water samples taken from a water purification system. There were $n = 18$ sequential samples which were taken and analyzed for the number of bacterial colonies, see Hoffmann (2003) for further details.

We obtained a p-value greater than 0.1, although the sample size was very small and only one zero count was reported. Both the Comp and DT methods gave similar estimates, whereas the S&S approach gave larger estimates and standard errors.

Migrating woodlark count data

Our fourth example uses count data collected on migrating woodlarks at Hanko bird observatory, during autumns 2007–2009. These data were previously analysed and obtained from Lindén & Mäntyniemi (2011).

We analysed each year separately. The S&S approach clearly overestimated both parameters for years 2007 and 2008, whereas the proposed DT method gave similar estimates to the Comp method for κ and μ. Only for the year 2009 were all methods similar, this was quite surprising since the sample sizes and sample means were similar to other years.

6.2 Case study 2: Barro Colorado Island tree abundance data

Our second case study uses a well-known data set consisting of tree abundance observations collected in a tropical forest plot on the Barro Colorado Island in Panama,

see `http://www.ctfs.si.edu`. The Barro Colorado Island has been protected by the Smithsonian Tropical Research Institution (STRI) since 1946, see Condict (1995) and Hwang & Shen (2010) for further details. The STRI established a 50-hectare ($500 \times 1,000$m) permanent plot to investigate dynamic changes within the forest. To date, the plot has been censused seven times in: 1981–1983, 1985, 1990, 1995, 2000, 2005 and 2010, where all free-standing trees (at least 1cm in diameter at breast height) were identified and located on a reference map.

We used the most abundance tree species *Hybanthus prunifolius* collected in the 1985 census, of which there were a total of 40,941 observed trees in the data. Firstly, we specified a strip quadrat to be of size 10×2m. As in Figure 1, each strip is surveyed from left to right, under the proportionate sampling assumption, such that the detection time is the shortest distance from the left boundary until an individual tree is observed. Within the forest region we collected 100 quadrats of this type where the bottom-left coordinate of each quadrat corresponds to $(100j, 50k)$ for $j = 1, \ldots, 10$ and $k = 1, \ldots, 10$. For each quadrat we also collected the complete count data where the average frequency was 1.46 with a sample s.d. of 1.68. The frequencies (m_0, m_1, m_2) used in S&S were $(39, 23, 38)$. The average of the first detection time (i.e., the percentage of surveyed area in a quadrat to detect the first tree) was 0.61 with a sample s.d. of 0.38.

As in Section 6.1, we fitted each model and give the results in the top-half of Table 2. First, we checked the goodness-of-fit using a Chi-square test (via the `fitdistrplus`-package) on the complete data; this yielded a p-value of 0.479 suggesting these count data are likely to follow a negative binomial distribution. We also checked the goodness-of-fit for the DT method where we obtained a p-value of 0.263 with 6 degrees of freedom. The complete MLE and DT method gave very similar estimates for κ and μ with only minor differences. The S&S approach underestimated κ and slightly overestimated μ.

We further investigated model performance by extending the quadrat size to 20×2m for the 100 quadrats. We present the results in the bottom-half of Table 2. Here, the Chi-square goodness-of-fit p-values were 0.492 for the complete data and 0.680 with the 6 degrees of freedom for the DT method. The average frequency of the complete

count data was 3.09 with a sample s.d. of 3.02; $(m_0, m_1, m_2) = (23, 14, 63)$; and the average of the first detection time was 0.45 with a sample s.d. of 0.37. Once again the DT method performed very well (since $\bar{x}$ is now much larger due to doubling the size of the quadrat), yielding similar estimates compared to the complete MLE.

7 Discussion

In this study, we proposed that the (first) time to detection in each quadrat should be recorded so that estimation of negative binomial parameters is feasible from occurrence data. This is an alternative approach to data augmentation by sampling until the second detection as in Solow & Smith (2010) or extending the model to include dependence between presences in adjacent quadrats (Hwang & Huggins, 2015). We showed that under what we term proportionate sampling, the resulting estimators improve on the Solow & Smith (2010) estimates for a range of values for negative binomial parameters κ and μ, and are also cheaper to obtain when sampling cost is related to the length of survey time. In addition to being cheaper, our method is less invasive, as less of the area needs be surveyed, which may be advantageous for some species, and better conservation strategies can be developed. We also note that when collecting data according to the approach of Solow & Smith (2010), one may of course record the first two detection times and use these data to conduct inferences based on the likelihood – doing this would of be more efficient than our proposed approach. One limitation of the proposed method is that optimization algorithms may sometimes yield undesirable results or fail to converge. For example, in the simulation studies we found that the proposed method occasionally yielded extremely large estimates for the shape index parameter κ. This situation was worse when the mean parameter μ was small and/or the shape parameter κ was large – i.e., whenever the ratio of the variance to the mean approached one. This phenomena was also observed in Cruyff & Van der Heijden (2008) and Böhning (2015) who considered the negative binomial model in a different setting. To resolve this difficulty, Cruyff & Van der Heijden (2008) suggested

considering alternatives to the negative binomial model.

We also investigated the effects of mis-identification. First, we found that when using either occurrence data (as in the proposed method) or the complete count data case, estimates can be biased in the presence of mis-identification. Interestingly, using the method of moments, we found that there is an attenuation effect for estimating the aggregation index κ when mis-identification probabilities varied from sample to sample (which is plausible in real applications). Despite this, we found that the effect on estimating κ is usually minor when the coefficient of variation of C_i, σ_c/c is small. Secondly, we developed two correction methods when information on mis-identification is available: the simple coefficient adjustment and the marginal likelihood approach. The former was developed from the complete data case, and worked well for both the Solow & Smith (2010) and proposed methods. However, a more detailed investigation is still required to examine the effects of mis-identification when using both methods. The formal marginal likelihood approach can work well, but this method requires specification of the true distribution of C_i and more computation effort is required. In practice, we recommend using the simple coefficient adjustment method as it generally worked well in our simulation study and only required information for c and σ_c^2. The simple coefficient adjustment is a one-step iteration method which cannot be fully iterated, otherwise, it yields a meaningless estimate at infinity.

Finally, in the current work we only investigated "false negatives" but it would also be interesting to examine "false positives". If available, a doubling-checking procedure (i.e., we double check our sample to make sure the correct identification is made) can be used to remove false positive errors. But further work still remains in finding a correction method to deal with false positives if such doubling-checking procedures are not available. As suggested by a referee, another interesting extension is building a detection time model associated with the zero-inflated negative binomial model; this allows for extra zero observations which are commonly observed in ecology and biology data. Although this extension appears feasible, both the implementation and performance warrant a further detailed study.

Acknowledgements

We would like to thank the Associate Editor and the reviewers for their helpful comments and constructive suggestions. This work was supported by the University of Melbourne and the Ministry of Science and Technology of Taiwan.

References

Alexander, N., Moyeed, R., & Stander, J. (2000). Spatial modelling of individual-level parasite counts using the negative binomial distribution. *Biostatistics*, **1**, 453–463.

Böhning, D. (2015). Power series mixtures and the ratio plot with applications to zero-truncated count distribution modelling. *Metron*, **73**, 201–216.

Cameron, A. C. & Trivedi, P. (1998). *Regression Analysis of Count Data*. Cambridge University Press, Cambridge.

Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective.* 2nd Ed. London: Chapman & Hall/CRC.

Condict, R. (1995). Research in large, long-term tropical forest plots. *Trends in Ecology and Evolution* **10**, 18–22.

Conlisk, E., Conlisk, J., & Harte, J. (2007). The impossibility of estimating a negative binomial clustering parameter from presence-absence data: A comment on He and Gaston. *American Naturalist*, **170**, 651–654.

Conlisk, E., Conlisk, J., Enquist, B., Thompson, J., & Harte, J. (2009). Improved abundance prediction from presence-absence data. *Global Ecology and Biogeography*, **18**, 1–10.

Cruyff, M. J. & Van der Heijden, P. G. (2008). Point and interval estimation of the population size using a zero-truncated negative binomial regression model. *Biometrical Journal*, **50**, 1035–1050.

Cruyff, M. J. & Van der Heijden, P. G. (2014). Sensitivity analysis and calibration of population size estimates obtained with the zero-truncated Poisson regression model. *Statistical Modelling*, **14**, 361–73.

Daley, D. J. & Vere-Jones, D. (2003). *An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods*. 2nd Ed. Springer.

Delignette-Muller, M. L. & Dutang, C. (2015). `fitdistrplus`: An `R` Package for Fitting Distributions. *Journal of Statistical Software*, **64**, 1–34.

Diggle, P. J. & Milne, R. K. (1983). Negative binomial quadrat counts and point processes. *Scandinavian Journal of Statistics*, **10**, 257–267.

Fromont, E., Morvilliers, L., Artois, M., & Pontier, D. (2001). Parasite richness and abundance in insular and mainland feral cats : insularity or density? *Parasitology*, **123**, 143–151.

Gregoire, G. (1983). Negative Binomial distributions for point processes. *Stochastic Processes and their Applications*, **16**, 179–188.

Hilborn, R. & Mangel, M. (1997). *The Ecological Detective. Confronting Models with data*. New Jersey: Princeton University Press.

Hoffmann, D. (2003). Negative binomial control limits for count data with extra-Poisson variation. *Pharmaceutical Statistics*, **2**, 127–132.

Hwang, W-H. & Huggins, R.M. (2015). Estimating abundance from presence-absence maps via a paired negative binomial model. *Scandinavian Journal of Statistics*, in press.

Hwang, W. H. & Shen, T. J. (2010). Small-sample estimation of species richness applied to forest communities. *Biometrics*, **66**, 1052–1060.

Lindén, A. & Mäntyniemi, S. (2011). Using the negative binomial distribution to model overdispersion in ecological count data. *Ecology*, **92**, 1414–1421.

Manly, B. F. J. & Navarro Alberto, J. A. (2015). *Introduction to Ecological Sampling.* London: Chapman & Hall/CRC.

Pielou, E. C. (1977). *Mathematical Ecology.* John Wiley, New York.

R Development Core Team. (2016). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. ISBN 3–900051–07–0. Available at `http://www.r-project.org`.

Solow, A. R. & Smith, W. K. (2010). On predicting abundance from occupancy. *American Naturalist*, **176**, 96–98.

Ver Hoef, J. M. and Boveng, P. L. (2007). Quasi-Poisson vs negative binomial regression: how should we model overdispersed count data? *Ecology*, **88**, 2766–2772.

Yin, D. & He, F. (2014). A simple method for estimating species abundance from occurrence maps. *Methods in Ecology and Evolution*, **5**, 336–343.

Winkelmann, R. (2008). *Econometric Analysis of Count Data: 5th Edition.* Springer, New York.

Appendix

Proof of Proposition 2

Proof. The expectation $E(T_1)$ can be derived from (2) directly, i.e., $E(T_1) = \int_0^1 g_{\boldsymbol{\theta}}(t)dt + p_0(\boldsymbol{\theta})$. Nevertheless, we give an alternative comprehensive proof. To derive $E(T_1)$, we show that $E(T_1 \mid X = x) = 1/(x+1)$ for all $x \geq 0$. The assertion is obviously true when $x = 0$. Given that $x > 0$, let U_j be the time to detect the jth individual for $j = 1, \ldots, x$, then $T_1 = U_{(1)}$ which is the minimum statistic of each U. Recall that X is equivalent to a gamma Poisson random variable. Write $E\{U_{(1)}\} = E[E\{U_{(1)} \mid \lambda\}]$ where $\lambda \sim Gamma(\kappa, \mu/\kappa)$ and $X \mid \lambda \sim Po(\lambda)$. Note that $U_j \mid \lambda$ are independent and identically uniform distributed and the jth order statistics $U_{(j)} \sim Beta(j, x-j+1)$, when λ is fixed. Therefore $E\{U_{(1)} \mid \lambda\} = 1/(1+x)$ which is a constant with respect to λ and so we have $E(T_1 \mid X = x) = E\{U_{(1)}\} = 1/(1+x)$.

For the second part of the proposition, note that $E(T_2 \mid X = x) = 1$ when $x = 0, 1$ and a similar argument to the above shows that $E(T_2 \mid X = x) = E\{U_{(2)}\} = 2/(x+1)$ when $x \geq 2$. Consequently, we have

$$\begin{aligned}
E(T_2) &= E(T_2 \mid X \geq 2)P(X \geq 2) + E(T_2 \mid X = 1)P(X = 1) + E(T_2 \mid X = 0)P(X = 0) \\
&= 2E(T_1 \mid X \geq 2)P(X \geq 2) + P(X = 1) + P(X = 0) \\
&= 2E(T_1 \mid X \geq 2)P(X \geq 2) + 2E(T_1 \mid X = 1)P(X = 1) + P(X = 0) \\
&= 2E(T_1) - P(X = 0).
\end{aligned}$$

Now,

$$E\left(\frac{1}{X+1}\right) = \sum_{x=0}^{\infty} \frac{\Gamma(\kappa+x)}{\Gamma(\kappa)(x+1)!} \frac{\kappa^\kappa \mu^x}{(\kappa+\mu)^{x+\kappa}},$$

which is equal to $\log(1+\mu)/\mu$ when $\kappa = 1$. For $\kappa \neq 1$, we have

$$\begin{aligned}
E\left(\frac{1}{X+1}\right) &= \frac{\kappa}{\mu(\kappa-1)} \left\{ \sum_{x=1}^{\infty} \frac{\Gamma(\kappa+x)}{\Gamma(\kappa-1)x!} \frac{\kappa^{\kappa-1}\mu^x}{(\kappa+\mu)^{x+\kappa-1}} \right\} \\
&= \frac{\kappa}{\mu(\kappa-1)} \left\{ 1 - \left(\frac{\kappa}{\kappa+\mu}\right)^{\kappa-1} \right\}.
\end{aligned}$$

As a consequence,

$$E(T_2 - T_1) = E\{1/(X+1)\} - P(X=0)$$
$$= \begin{cases} \dfrac{\kappa}{\mu(\kappa-1)}\left\{1 - \left(\dfrac{\kappa}{\kappa+\mu}\right)^{\kappa-1}\right\} - \left(\dfrac{\kappa}{\kappa+\mu}\right)^{\kappa}, & \kappa \neq 1 \\ \dfrac{\log(1+\mu)}{\mu} - \dfrac{1}{1+\mu} & \kappa = 1. \end{cases}$$

□

Proof of Proposition 3

Proof. The proof is straightforward once we note that as $T_1 \mid x \sim Beta(1, x)$ when $x > 0$, then we have

$$E(T_1^2 \mid X = x) = \frac{2}{(X+2)(X+1)}.$$

Similarly, we have

$$E(T_2^2 \mid X = x) = \frac{6}{(x+2)(x+1)},$$

when $x > 0$ so that

$$E(T_2^2) = 6E\left\{\frac{1}{(X+2)(X+1)}\right\} - 2P(X=0).$$

Next, notice that $E(T_1T_2 \mid x) = 1$ for $x = 0$, $E(T_1T_2 \mid x) = E(T_1 \mid x) = 1/2$ for $x = 1$, and $E(T_1T_2 \mid x) = \dfrac{3}{(x+2)(x+1)}$ by noting that (T_1, T_2) are identically distributed to the first two order statistics from a set of x random uniform variables. It turns out that

$$E(T_1T_2) = 3E\left\{\frac{1}{(X+2)(X+1)}\right\} - \frac{P(X=0)}{2},$$

which completes the proof. □

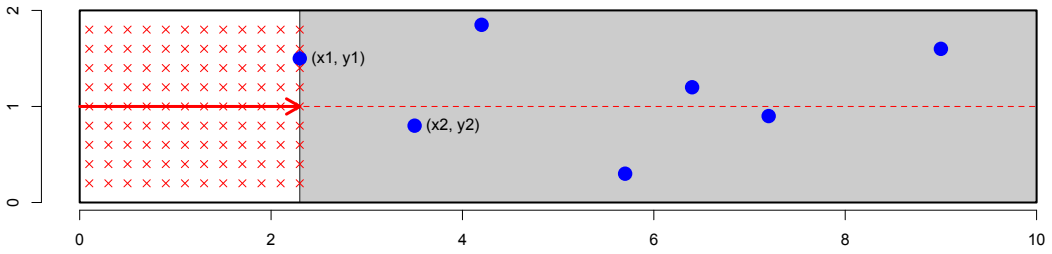

Figure 1: *The schematic above demonstrates how the "first detection time" is recorded for a given strip quadrat of size* $10 \times 2m$. *Suppose there are 7 trees (blue dots) in the strip quadrat and we conduct strip transect sampling from left to right. As soon as we detect a tree we record the tree's location coordinates* (x_1, y_1) *and stop sampling for this quadrat. The red crosses above display how much area has been sampled before the first tree is detected. Suppose that the required time to survey a region is proportional to its area, then the detection time can be specified as* $x_1/10$, *i.e., it is the percentage of survey effort to complete the census of the strip quadrat. Note that the Solow & Smith (2010) approach stops at the second tree with coordinates* (x_2, y_2), *requiring additional sampling effort.*

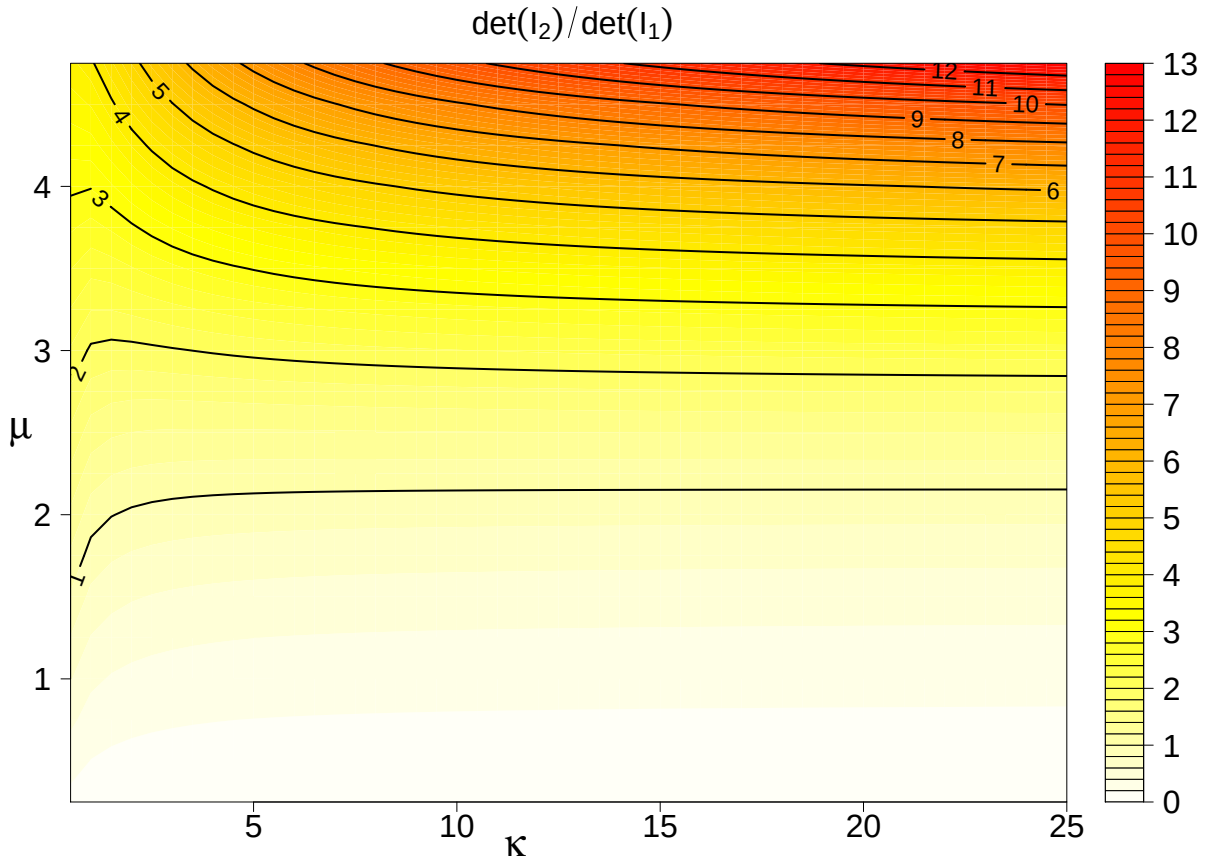

Figure 2: *An image plot showing the ratio of* $det(I_2(\boldsymbol{\theta}))/det(I_1(\boldsymbol{\theta}))$*, see text for details on notation. Numerical integration was used to calculate the expected Fisher information. Notice that the proposed model is generally more efficient when* $\mu > 2$.

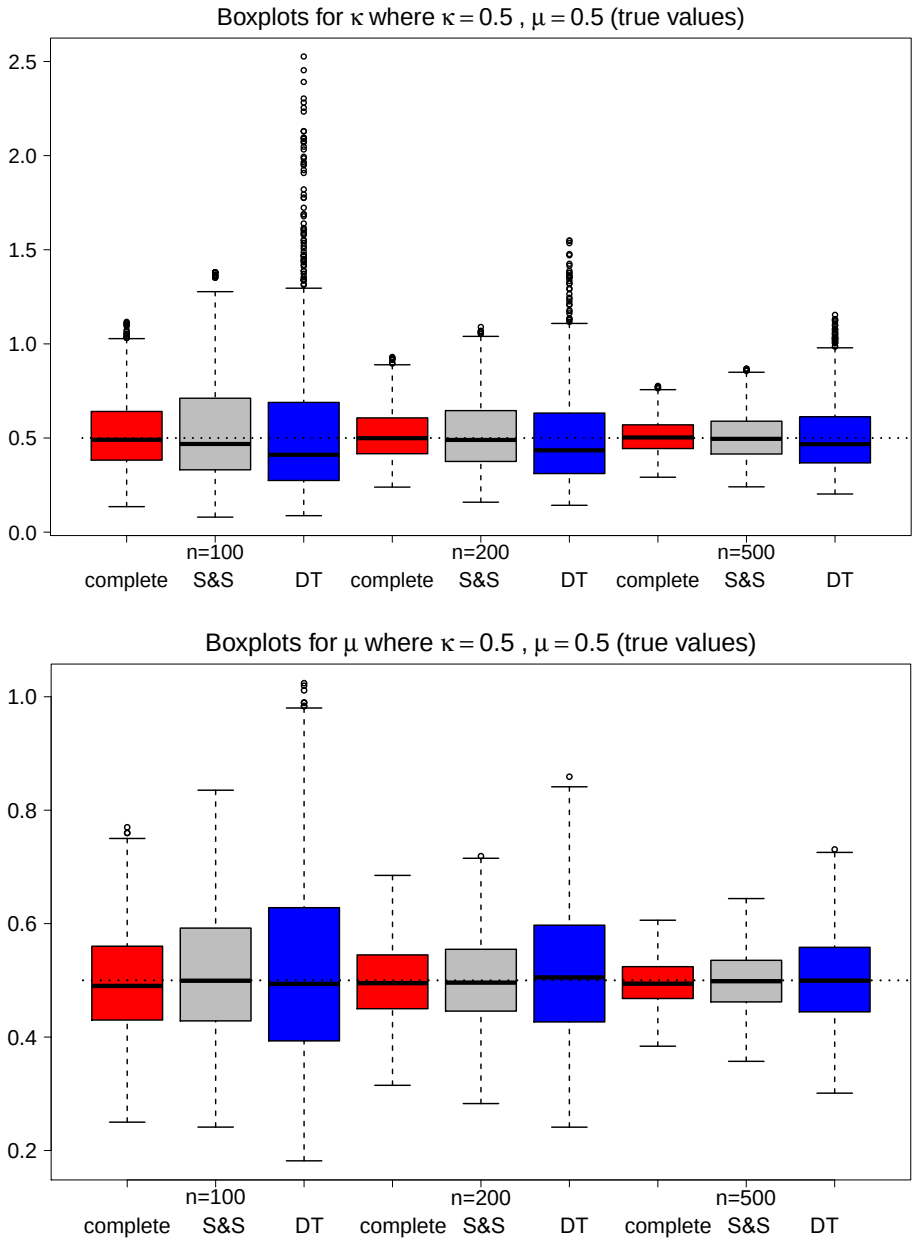

Figure 3: *Boxplot for estimates of κ (top) and μ (bottom) when $\kappa = 0.5$ and $\mu = 0.5$ for each sample size for simulation study 1.*

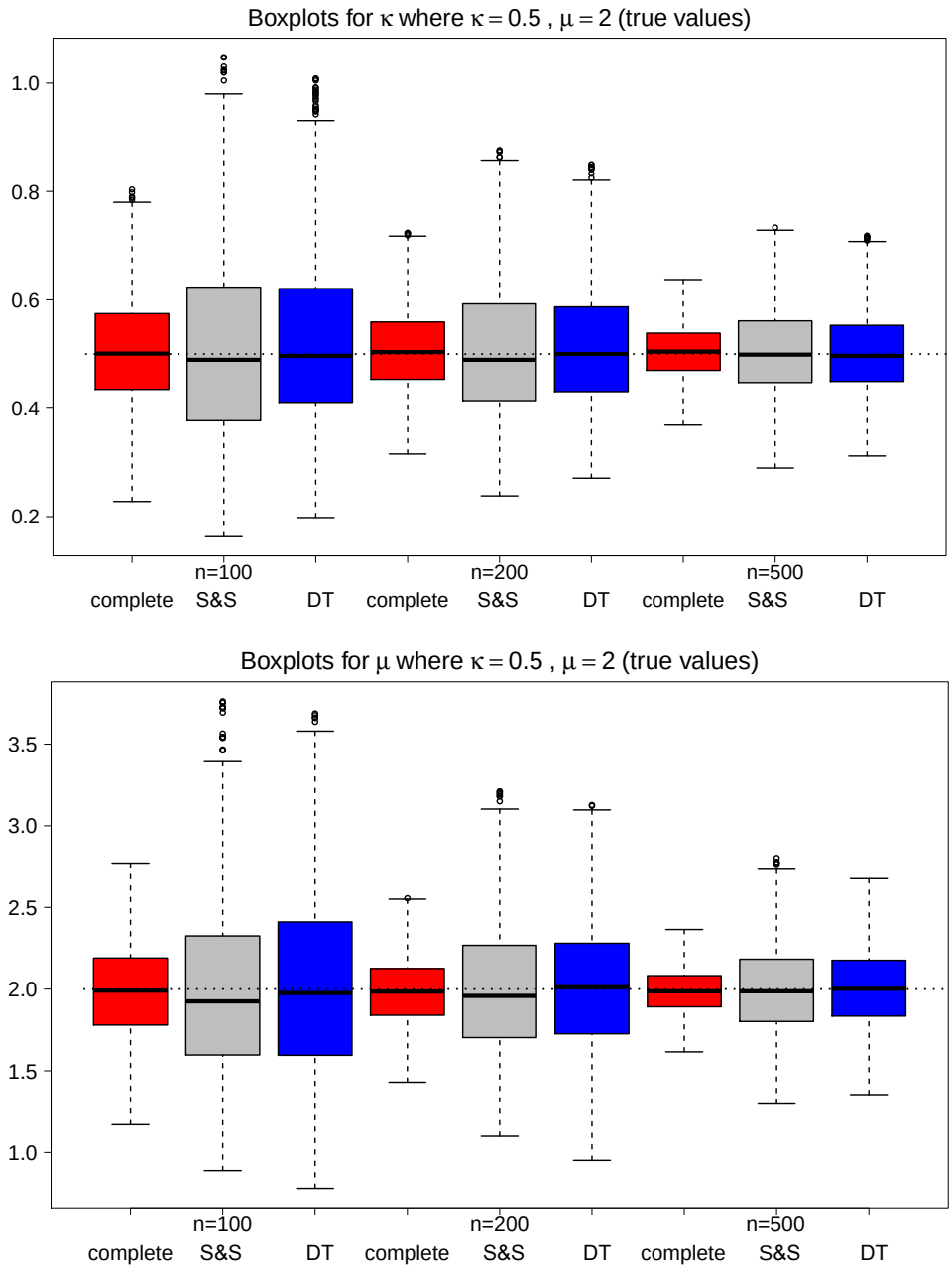

Figure 4: *Boxplot for estimates of κ (top) and μ (bottom) when $\kappa = 0.5$ and $\mu = 2$ for each sample size for simulation study 1.*

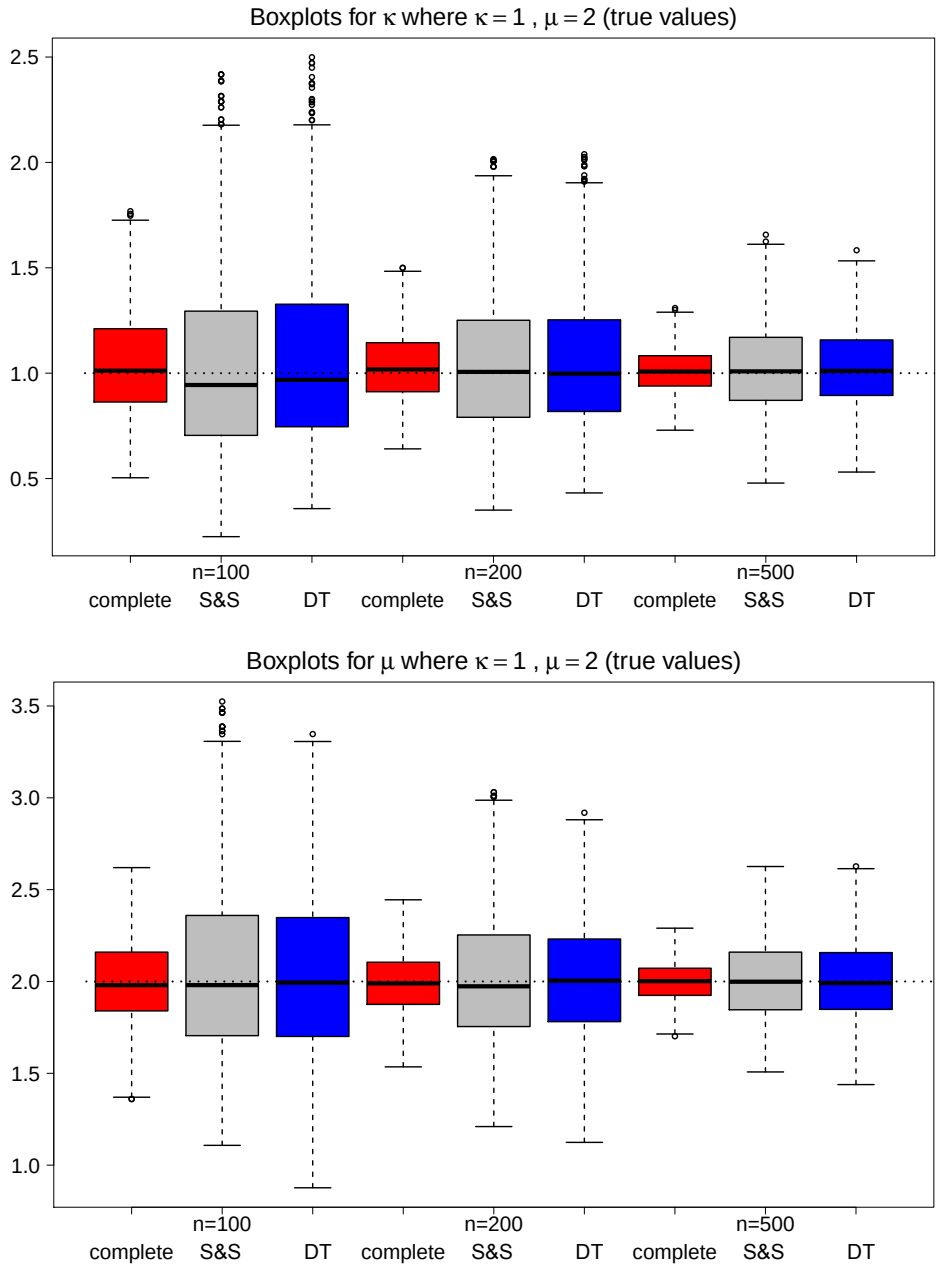

Figure 5: *Boxplot for estimates of κ (top) and μ (bottom) when $\kappa = 1$ and $\mu = 2$ for each sample size for simulation study 1.*

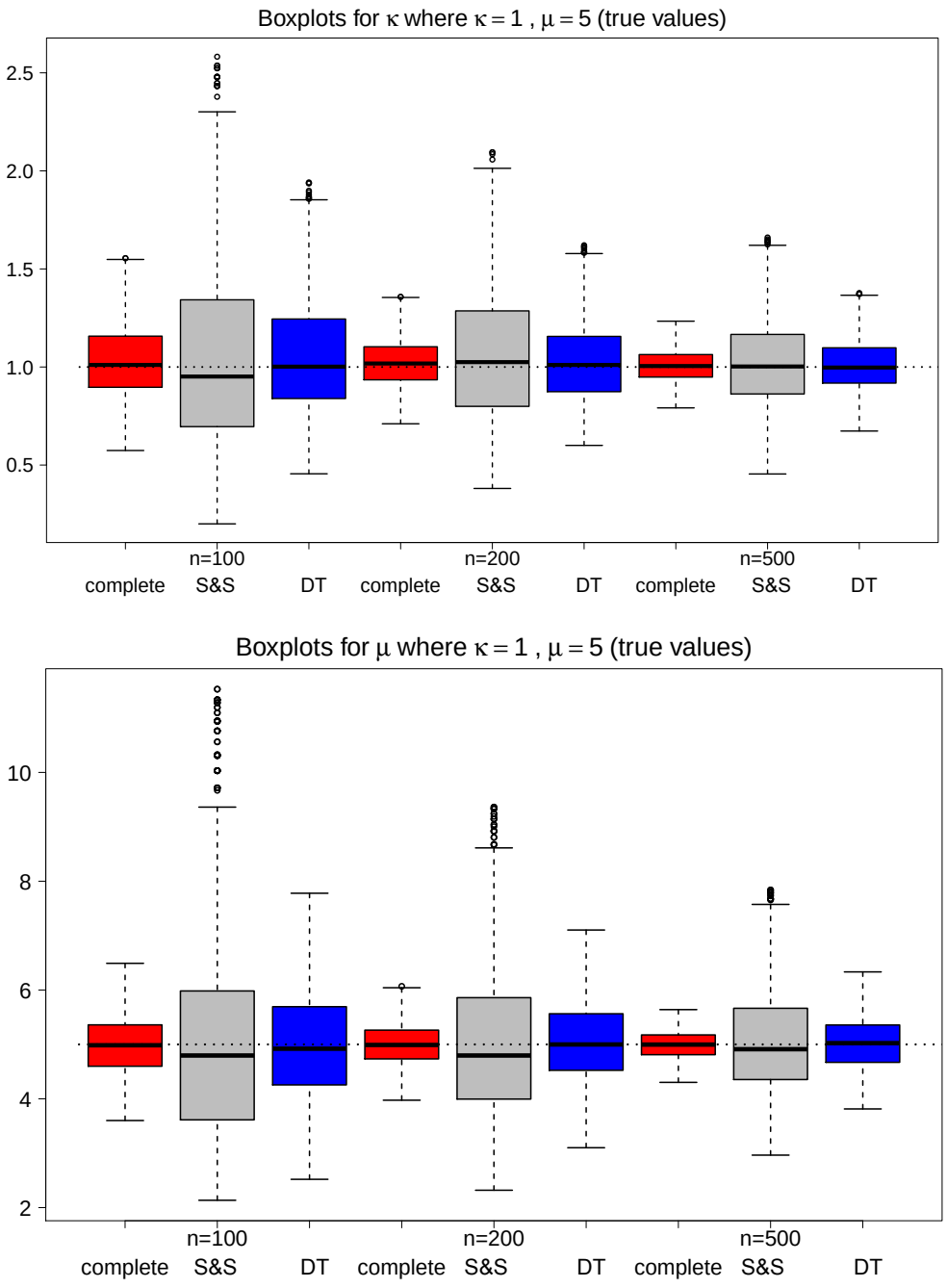

Figure 6: *Boxplot for estimates of κ (top) and μ (bottom) when $\kappa = 1$ and $\mu = 5$ for each sample size for simulation study 1.*

Table 1: Parameter estimates (with standard error estimates in parentheses) for the complete (Comp) MLE, the (S&S, Solow & Smith, 2010) approach, and the proposed detection time (DT) method for each example of case study 1 (Section 6.1). Here, for the DT estimates we denote DT_a and DT_m as the sample average and median, respectively, from the 1000 simulated data sets. We also give the average (and s.d.) for the detection times for each method. See text for additional details.

Parasite abundance					
$n_0 = 14$	para.	Comp (s.e.)	S&S (s.e.)	DT_a (ave s.e.)	DT_m (med s.e.)
$\bar{x} = 8.68$	κ	0.397 (0.083)	0.659 (0.389)	0.473 (0.149)	0.462 (0.141)
$sd(x) = 14.29$	μ	8.681 (1.936)	4.313 (2.650)	7.922 (3.123)	7.436 (2.902)
p-value = 0.11	time	1.0 (0.0)	0.584 (0.017)	0.424 (0.022)	
Albatross abundance					
$n_0 = 807$	para.	Comp (s.e.)	S&S (s.e.)	DT_a (ave s.e.)	DT_m (med s.e.)
$\bar{x} = 0.28$	κ	0.062 (0.009)	0.053 (0.011)	0.065 (0.017)	0.064 (0.016)
$sd(x) = 1.25$	μ	0.279 (0.041)	0.336 (0.087)	0.280 (0.072)	0.271 (0.068)
p-value = 0.098	time	1.0 (0.0)	0.971 (0.002)	0.935 (0.003)	
Bacterial abundance					
$n_0 = 1$	para.	Comp (s.e.)	S&S (s.e.)	DT_a (ave s.e.)	DT_m (med s.e.)
$\bar{x} = 10.61$	κ	1.265 (0.474)	1.071 (1.830)	1.411 (1.010)	1.305 (0.822)
$sd(x) = 8.87$	μ	10.609 (2.351)	14.839 (39.146)	11.873 (5.824)	10.625 (4.879)
p-value = 0.19	time	1.0 (0.0)	0.355 (0.031)	0.205 (0.033)	
Migrating woodlark abundance, Year 07					
$n_0 = 39$	para.	Comp (s.e.)	S&S (s.e.)	DT_a (ave s.e.)	DT_m (med s.e.)
$\bar{x} = 4.55$	κ	0.184 (0.041)	0.103 (0.055)	0.194 (0.057)	0.192 (0.055)
$sd(x) = 11.21$	μ	4.551 (1.283)	34.735 (79.981)	4.642 (2.214)	4.228 (1.953)
p-value = 0.82	time	1.0 (0.0)	0.749 (0.013)	0.649 (0.015)	
Migrating woodlark abundance, Year 08					
$n_0 = 38$	para.	Comp (s.e.)	S&S (s.e.)	DT_a (ave s.e.)	DT_m (med s.e.)
$\bar{x} = 5.08$	κ	0.202 (0.045)	0.133 (0.065)	0.177 (0.046)	0.174 (0.045)
$sd(x) = 9.23$	μ	5.086 (1.368)	14.584 (23.519)	6.784 (3.121)	6.502 (2.920)
p-value = 0.30	time	1.0 (0.0)	0.701 (0.011)	0.620 (0.013)	
Migrating woodlark abundance, Year 09					
$n_0 = 39$	para.	Comp (s.e.)	S&S (s.e.)	DT_a (ave s.e.)	DT_m (med s.e.)
$\bar{x} = 3.06$	κ	0.222 (0.052)	0.219 (0.100)	0.236 (0.079)	0.229 (0.072)
$sd(x) = 6.65$	μ	3.055 (0.796)	3.148 (2.293)	3.211 (1.453)	2.909 (1.278)
p-value = 0.81	time	1.0 (0.0)	0.780 (0.013)	0.665 (0.017)	

Table 2: Parameter estimates (with standard error estimates in parentheses) for the complete (Comp) MLE, the (S&S, Solow & Smith, 2010) approach, and the proposed detection time (DT) method for case study 2 (Section 6.2) using the tree species *Hybanthus prunifolius* data under two different quadrat sizes: $10 \times 2m$ (top-half) and $20 \times 2m$ (bottom-half). The sample size is 100.

10×2m, $n_0 = 39$, $n_1 = 23$, $\bar{x} = 1.46$, $sd(x) = 1.68$			
para.	Comp (s.e.)	S&S (s.e.)	DT (s.e.)
κ	1.315 (0.426)	0.921 (0.452)	1.244 (0.779)
μ	1.460 (0.175)	1.639 (0.382)	1.411 (0.350)
20×2m, $n_0 = 23$, $n_1 = 14$, $\bar{x} = 3.09$, $sd(x) = 3.02$			
para.	Comp (s.e.)	S&S (s.e.)	DT (s.e.)
κ	1.344 (0.310)	0.691 (0.306)	1.240 (0.486)
μ	3.089 (0.319)	5.104 (2.508)	2.854 (0.621)

Author/s:
Hwang, W-H;Huggins, R;Stoklosa, J

Title:
Estimating negative binomial parameters from occurrence data with detection times

Date:
2016-11-01

Citation:
Hwang, W. -H., Huggins, R. & Stoklosa, J. (2016). Estimating negative binomial parameters from occurrence data with detection times. BIOMETRICAL JOURNAL, 58 (6), pp.1409-1427. https://doi.org/10.1002/bimj.201500239.